Evaluation of privacy in high dynamic range video sequences

Martin Řeřábek[a], Lin Yuan[a], Lukáš Krasula[b], Pavel Korshunov[a], Karel Fliegel[b], and Touradj Ebrahimi[a]

[a]Multimedia Signal Processing Group (MMSPG), Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland; [b]Multimedia Technology Group (MMTG), Czech Technical University in Prague (CTU in Prague), Czech Republic

ABSTRACT

The ability of high dynamic range (HDR) to capture details in environments with high contrast has a significant impact on privacy in video surveillance. However, the extent to which HDR imaging affects privacy, when compared to a typical low dynamic range (LDR) imaging, is neither well studied nor well understood. To achieve such an objective, a suitable dataset of images and video sequences is needed. Therefore, we have created a publicly available dataset of HDR video for privacy evaluation PEViD-HDR, which is an HDR extension of an existing Privacy Evaluation Video Dataset (PEViD). PEViD-HDR video dataset can help in the evaluations of privacy protection tools, as well as for showing the importance of HDR imaging in video surveillance applications and its influence on the privacy-intelligibility trade-off. We conducted a preliminary subjective experiment demonstrating the usability of the created dataset for evaluation of privacy issues in video. The results confirm that a tone-mapped HDR video contains more privacy sensitive information and details compared to a typical LDR video.

Keywords: High Dynamic Range, dataset, video surveillance, privacy evaluation

1. INTRODUCTION

The widespread usage of digital video surveillance systems has increased the concerns for privacy violations. Many media reports indicate that the number of people caught on camera by video surveillance systems is rapidly increasing, and it seems this number will continue to increase in the foreseeable future. Such recording of personal data combined with video analytics and ease of access to personal information through social networks give rise to an unprecedented ability to collect privacy sensitive content on individuals. The advances of new capturing and display technologies such as ultra high definition (UHD) or high dynamic range (HDR) increase privacy concerns even more.

HDR imaging can represent more accurately, compared to standard low range (LDR) imaging, the range of intensity levels found in real scenes, from direct sunlight to faint starlight. The most commonly used technique for capturing HDR images is called *image fusion*. Using this technique, an HDR image can be obtained by combining several images (called *brackets*) of the same scene shot at different exposures to capture the necessary dynamic range. This approach allows to circumvent the limited dynamic range of the conventional camera sensors, making HDR a popular alternative to conventional LDR.[1]

Further author information (send correspondence to Martin Řeřábek):
Martin Řeřábek: martin.rerabek@epfl.ch
Lin Yuan: lin.yuan@epfl.ch
Lukáš Krasula: krasuluk@fel.cvut.cz
Pavel Korshunov: pavel.korshunov@epfl.ch
Karel Fliegel: fliegek@fel.cvut.cz
Touradj Ebrahimi: touradj.ebrahimi@epfl.ch

The ability of HDR to capture higher details and contrasts raises concerns regarding more invasive privacy intrusions. Private life activities or facts can be more easily exposed now to common view due to equally high quality details in dark and bright image regions offered by the new technology. For example, it becomes easier to use video cameras intended for outdoor surveillance (bright day outside) to maliciously spy on people inside buildings (dark inside).

It is important to study and analyze this issue and for that a dataset is necessary. Currently available HDR datasets are commonly used for evaluation of tone-mapping algorithms or appeal of HDR compared to LDR images and are not suitable for studying the effects of HDR on privacy in video, since they do not normally have people in them. One of the main reasons for the lack of people or any naturally moving objects in HDR images or video datasets is the immaturity of the HDR capturing technology and the complexity of shooting high quality HDR video content. There are also many non-HDR datasets for evaluation of video analytics (detection, recognition, and tracking algorithms), such as FERET dataset, Labeled Faces in the Wild (LWF), PETS 2007, etc. But since these datasets were not designed with privacy issues in mind, they are not suitable for evaluation of privacy related aspects. To the authors knowledge, only a couple of datasets for evaluation of privacy exists, one is PEViD,[2] which is not an HDR dataset, and another is a dataset of HDR images for privacy,[3] which does not contain video content and hence is not suitable for evaluation of typical video surveillance scenes.

To compensate the lack of the HDR video datasets for evaluation of privacy, we created a publicly available PEViD-HDR video dataset[*] that contains 68 video sequences. The sequences depict different surveillance indoor and outdoor scenarios shot under highly variable lighting conditions, including deep shades, sunny outdoor, and dark indoor scenes. Video sequences were shot by three HDR capable video cameras: Bosch DINION IP dynamic 7000 HD (NBN-932V-IP), JAI AD-132 GE, and BlackMagic Pocket Cinema Camera (BMPCC). To demonstrate the usefulness of such video dataset for the evaluation of privacy issues in HDR video, we conducted a preliminary subjective experiment using 24 naïve subjects. All raw video data from the cameras and evaluation data is included into the dataset for research use.

Therefore, the following summarizes the main contributions of this paper:

- HDR video dataset designed for privacy evaluations;
- Video obtained by three state of the art HDR video cameras including Bosch DINION IP dynamic 7000 HD, the IP industrial HDR capable surveillance camera;
- Results of the subjective evaluation of privacy issues in HDR video in the dataset.

In the rest of the paper, after presenting the related work, we describe in details PEViD-HDR video dataset for privacy evaluations in HDR video, including the scenarios and capture process, followed by the presentation of subjective evaluation and discussion of the results on how much HDR video is more privacy intrusive when compared to typical LDR video.

2. BACKGROUND AND RELATED WORK

In this section, an overview of related work in areas of privacy protection and HDR imaging is provided.

2.1 Privacy

A lot of research effort in the past was focused on approaches to incorporate privacy protection into existing security systems and frameworks, typically via implementing access rights management and policies.[4–7] Another large body of work is on development of algorithms and methods to protect visual privacy, such as using watermarking to hide visual personal information,[8] scrambling techniques to reversibly distort privacy sensitive regions,[9] removal of unauthorized personnel from the video feed,[10] encoder independent geometrical-based reversible distortions.[11,12] However, little was done towards understanding of privacy issues in practical multimedia applications, analysis of the effect of privacy protection on the main system utility, and development of effective evaluation methodologies that take into account both the context and content.

Some work has been done for the evaluation of privacy filters (tools that protect visual privacy) objectively. The objective evaluation of several primitive privacy filters was first performed by Newton et al.,[13] where the

[*] http://mmspg.epfl.ch/pevid-hdr

authors demonstrated that such filters cannot adequately protect from the successful face recognition, because recognition algorithms are robust. The robustness of face recognition and detection algorithms to primitive distortions is also reported.[14] Further, in work by Dufaux et al.,[15] a framework is defined to evaluate the performance of face recognition algorithms applied to images altered by various obfuscation methods. Another study[16] considered an automated video surveillance system and focused on finding the privacy-intelligibility tradeoff using objective metrics such as face detection and face recognition.

However, since privacy is a subjective notion, its evaluation should be done subjectively.[17] The authors considered the problem of finding the balance between the ability of human guards to perform a surveillance task and adequate protection of privacy. A subjective methodology and protocol were defined for evaluation of privacy protection tools, focusing on two important aspects: (i) how much of the privacy is protected by such a tool and (ii) how much it impacts the efficiency of the underlying surveillance task (intelligibility). The pixelization filter showed the best performance in terms of balancing between privacy protection and allowing sufficient intelligibility. Masking filter, instead, demonstrated the highest privacy protection with low incorrectness and high uncertainty, which can be suitable for the higher security surveillance applications.

2.2 HDR imaging

High dynamic range (HDR) imaging is expected, together with ultra-high definition (UHD) and high frame rate (HFR) video, to become a technology that may change photo, TV, and film industries. Many different subjective evaluations have been previously performed to compare different tone-mapping operators for HDR images and video. Main focus of these studies was either on determining a more superior approach to tone-mapping or establishing an evaluation methodology for subjective evaluation of HDR content. One of the first subjective evaluations of HDR images was performed by Ledda et al.[18] The authors used paired comparison to evaluate the perceptual quality of six different tone-mapping algorithms. An HDR display was used as reference display for 48 subjects. The focus of this work was on the evaluation methodology for the subjective comparison of HDR images in a controlled environment. The evaluations provided the performance ranking of different tone-mapping algorithms leading to different perceptual qualities in color and gray images. Similar studies were conducted to determine the appeal of HDRi,[19] usefulness of HDR for astronomical images,[20] how accurately tone-mapping algorithms represent reality,[21] on objective metrics of HDR,[22] and on using HDR for 3D content.[23]

However, as opposed to the above work, the goal of our study is not to find the best tone-mapping algorithm but to demonstrate the invasive nature of HDR imaging, specifically HDR video, which threatens privacy of people. For this purpose, we need a comprehensive dataset of HDR video showing people in different video surveillance scenarios, which together with an effective evaluation methodology would allow us to test an HDR video for privacy intrusiveness in practical settings.

3. DATASET DESCRIPTION

Evaluation of privacy intrusiveness in HDR images requires a dataset containing in particular various people with a visible privacy-related information. There are many datasets for evaluation of video analytics, such as detection, recognition, and tracking algorithms. Also, there are HDR datasets commonly used for evaluation of tone-mapping algorithms or appeal of HDR compared to LDR. However, since these datasets were not designed with privacy issues in mind, they are not suitable for privacy related aspects evaluation. To the authors knowledge, there is one non-HDR video dataset for evaluation of privacy called PEViD[2] and one dataset of HDR images for privacy evaluations,[3] which does not contain video content and hence is not suitable for evaluation of privacy in video surveillance.

In the recent work by the authors,[2] the design principles of datasets suitable for evaluation of video analytics in video surveillance and datasets for evaluation of visual privacy protection tools were defined. Such a dataset should, on one hand, contain typical, as well as broad range of practical, surveillance scenarios with wide variety of privacy-related visual information, such as facial information, gender, personal identifiable items and accessories. On the other hand, it should bring out the specifics of HDR imaging, i.e., dynamic range of the content should be high enough for HDR to manifest itself. Following these principles, we have created an HDR video dataset for privacy evaluation as an extension of existing PEViD dataset.

3.1 Dataset creation

For our dataset, several typical indoor and outdoor video surveillance scenes with high contrast variety were considered and designed. In each of them, one to five participants perform different tasks, such as simple walking, stealing, fighting, exchanging items, riding a bicycle, standing, etc. Special effort was made to ensure the variety of gender, race, and different personal accessories that people carried or wore. All participants in dataset recordings were students and employees from Czech Technical university in Prague (CTU) and Ecole Polytechnique Fédérale de Lausanne (EPFL). Each participant of the recordings read and signed a disclaimer, which allows the dataset to be used for research purposes. The description of individual recorded indoor and outdoor scenes is shown in Table 1 and Table 2, respectively.

The HDR video data was recorded in two phases. At first, several indoor scenes exploiting the artificial lighting environment and one night scene was recorded at CTU. To obtain HDR video sequences, two low-cost state-of-the-art HDR cameras, JAI AD-132 GE[1]† and BlackMagic Pocket Cinema Camera‡ were used. In the second phase, only outdoor scenes were acquired in environment of EPFL by using, in addition to the above two cameras, Bosch DINION IP dynamic 7000 HD§ professional security HDR camera.

Bosch camera exploits proprietary algorithms to achieve up to 90 dB dynamic range. By using ONVIF standard¶, Bosch camera provides full HD video IP streaming with 30 fps frame rate. The output of Bosch camera is already tone-mapped to 8bit scale so it can be displayed on LDR monitor without further processing.

Camera JAI AD-132 GE is an imaging system equipped with two independent CCD sensors with Bayer masks and resolution of 1296×966 pixels allowing to capture HDR video sequences with 30 fps with a dynamic range of 20 f-stops. The camera enables to capture the RAW output of each of the sensors (CAM0 and CAM1) directly, or it composes the final HDR image based on one of the three possible internal HDR modes. However, these modes do not provide the highest dynamic range extension and the exact processing of the information from the particular sensors is unknown. Therefore, for this dataset, the individual output of each CCD was captured and the EV of each sensor was set in a way to exploit the maximum dynamic range of recorded scene.

Both Bosch and JAI cameras are connected to a PC via a special frame grabber‖, which enables the real time viewing of the data stream. For video recording, a specific multiple camera software Streampix** ensuring the synchronization of inputs from all cameras image sensors was used.

BMPCC uses only one full HD CMOS chip and enables data acquisition with two different settings. The first setting shoots video sequence in lossless DNG RAW format providing 13 f-stops of the dynamic range. The second setting records the ProRes 422 compressed format compliant to REC709 standard for high definition video. We used first set of settings to record videos for PEViD-HDR video dataset.

Table 1 and Table 2 show the Dynamic Range (DR) in dB computed for each sequence in the dataset as an average of DRs of ten consecutive frames, since lighting conditions did not change during the sequence recording. DR for each frame of BMPCC camera was determined as a ratio between 0.01 ($P_{0.01}L$) and 99.9 ($P_{99.9}L$) percentile of its luminance. For JAI camera, we calculated the DR value as follows:

$$DR_{JAI} = \frac{P_{99.9}L_{CCD_1}}{P_{0.01}L_{CCD_0}} \times \frac{T_{CCD_0}}{T_{CCD_1}}, \tag{1}$$

where $P_{0.01}L_{CCD_0}$ is a 0.01 percentile of luminance obtained from sensor with high exposition setting, $P_{99.9}L_{CCD_1}$ is a 99.9 percentile of luminance from sensor with low exposition setting, and T_{CCD_0} and T_{CCD_1} are exposition times of both CCD sensors. The sample frames of acquired video sequences representing recorded scenes are shown in Figure 1.

† http://www.jai.com/en/products/ad-132ge

‡ https://www.blackmagicdesign.com/products/blackmagicpocketcinemacamera

§ http://resource.boschsecurity.com/documents/Data_sheet_enUS_7720687371.pdf

¶ http://www.onvif.org/

‖ http://sine.ni.com/nips/cds/view/p/lang/cs/nid/205960

** http://norpix.com/products/streampix/streampix.php

Scene	Description	DR [dB]	Camera	Sequences	Length [s]
Corridor	Example of corridor video surveillance. Five men walking from dark to bright. Various accessories (bag, scarfs, glasses) and surveillance events (walking, fighting, stealing).	75	JAI	2	≈ 30
Room1	A theft attempt of a purse. A man and a woman walking and fighting when a man tries to steal a purse of a woman. Various accessories (hat, purse, bag, hood).	88	JAI	2	≈ 12
		94	BMPCC	2	
Room2	A man walking from dark to bright area where he picks/steals an item (pen/cellphone). Various accessories (hat, bag, pen, cell phone).	88	JAI	2	≈ 12
		94	BMPCC	2	
Studio	Three persons with various gender standing in bright spot. Behind them, in dark, the surveillance events, such as stealing, walking, smuggling, and fighting happen. Various accessories (bag, can, scarf, glasses, purse).	52	JAI	4	≈ 15
		87	BMPCC	4	
Night	A night video shooting containing two women and three men walking from bright to dark. Various accessories (bag, scarf, hat).	70	BMPCC	3	≈ 55

Table 1: Summary of the indoor and night scenes within PEViD-HDR dataset.

Scene	Description	DR [dB]	Camera	Sequences	Length [s]
Entrance1	Shot of glassy entrance of the building with several people and various activities (walking, fighting, standing, stealing) in each sequence. Cameras placed in shadow, sequences contain shear of the building shadow. Various items and accessories (mobile phone, bicycle).	-	Bosch	6	30-60
		59	JAI	6	
		61	BMPCC	2	
Entrance2	Shot from bright outdoor into the dark corridor in the entrance of the building. Cameras placed on the bright spot. Various activity in the entry hall as well as outside, such as fighting, walking, and stealing. Various items and accessories (hat, mobile phone, car, bag).	-	Bosch	3	20-40
		69	JAI	3	
Parking1	Shot on the underground parking garage containing various activity, such as walking, stealing, fighting, riding a bicycle. Various items and accessories (bottle, bag). Cameras placed in gentle shadow.	-	Bosch	4	25-40
		67	JAI	4	
		63	BMPCC	4	
Parking2	The same parking entrance as in previous scene. Camera placed inside of the parking entrance. Two people walking towards the entrance from bright area, one is ridding a motorbike to the garage.	-	Bosch	2	≈ 25
		73	JAI	2	
		72	BMPCC	2	
Tunnel	Girl walking from bright into dark tunnel. Guy waiting in the shadow trying to steal her bag. Two men standing in from of the tunnel on bright spot and talking.	-	Bosch	3	25-35
		70	JAI	3	
		53	BMPCC	3	

Table 2: Summary of the outdoor scenes within PEViD-HDR dataset. According to the provider, the maximum DR which can be recorder by Bosch camera is 90 dB. The actual DR of the scenes obtained by Bosch camera is unknown, since it only provides already tone-mapped 8 bit video stream.

(a) Corridor (b) Entrance1

(c) Room1 (d) Entrance2

(e) Room2 (f) Parking1

(g) Studio (h) Parking2

(i) Night (j) Tunnel

Figure 1: Examples of scenes recorded indoor and at night by JAI (a), and BMPCC camera (c),(e),(g),(i), and outdoor by Bosch camera (b),(d),(f),(h),(j).

4. EVALUATION EXPERIMENT

A subset containing six video sequences of PEViD-HDR video dataset for each camera was selected for subjective evaluation. Sequences were selected to represent both typical bright-to-dark (4 sequences) and dark-to-bright (2 sequences) scenarios. In the bright-to-dark scenarios, it is assumed that a surveillance camera monitors activity in bright areas of the scene (i.e., camera is tuned for the outdoor surveillance), while in the dark-to-bright scenarios, monitoring of the dark areas in the scene is assumed (i.e., camera is tuned for the indoor surveillance). Each video sequence was trimmed to be exactly ten seconds long. The body silhouette of one specific person was manually annotated in each video with ViPER-GT annotation tool*. The selected subset was then preprocessed to create HDR video versions, which were tone-mapped so that they can be displayed on the standard LDR monitor for the subjective evaluations. The examples of frames from selected video sequences for all acquisition systems used in our experiment are shown in Figure 2. All details about selected data preprocessing including tone-mapping are described in the next subsection.

4.1 Preprocessing of video sequences

Since HDR modes of JAI HDR camera do not provide the highest dynamic range extension and the exact processing of the information from the particular sensors is unknown, the outputs of the two independent CCDs were used to create HDR video. The 12-bit output of each chip is saved in TIFF format (16-bit integer). Firstly, both of the outputs were preprocessed to be in the same domain. The normalization was done as

$$I_{CCD1,norm} = \frac{I_{CCD1} - min[I_{CCD1}, I_{CCD2}]}{(2^{16}-1) - min[I_{CCD1}, I_{CCD2}]}, I_{CCD2,norm} = \frac{I_{CCD2} - min[I_{CCD1}, I_{CCD2}]}{(2^{16}-1) - min[I_{CCD1}, I_{CCD2}]}, \qquad (2)$$

where I_{CCD1} and I_{CCD2} are the frames obtained from the first and the second CCD, respectively. The global minimum is subtracted because of possible shift of dynamic range during mapping the 12-bit output in the 16-bits. The assumption here is that one of the exposure times is very short and thus the image with under-exposed areas should contain some minimal possible values. Both of the normalized images were then divided by the respective exposure times used during their capturing. The final HDR frame was obtained by averaging the two results. For visualization, iCAM06[24] TMO was used.

As mentioned above, the BMPCC offers dynamic range of 13 f-stops and allows capturing into lossless CinemaDNG RAW†† with resolution of 1920×1080 pixels. The captured RAW image files were at first losslessly converted into TIFF sequences (16-bit per color channel) using dcraw utility‡‡. The obtained TIFF sequences were then processed using two approaches:

A - The whole 13 f-stops extended dynamic range EDR was preserved and iCAM06[24] TMO was applied for visualization.

B - The capturing process with LDR camera was simulated.

In the latter case, the EDR of the linear RAW image data was clipped using simple histogram stretching procedure with highlights or shadows preserved for bright-to-dark or dark-to-bright scenarios, respectively. The simulated DR was set to 8 f-stops (approx. 48 dB), which is a typical DR for analog security CCTV cameras still in use. The pixel values clipped to 8 f-stops were then transformed into 8-bit per color channel representation using a simple global TMO (linear tone-mapping with inverse gamma correction, $\gamma = 2.2$) for visualization.

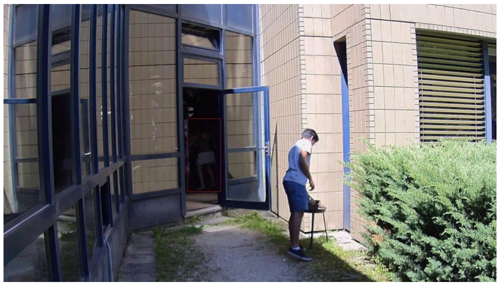
All LDR and HDR video sequences were further annotated in order to use them for evaluation of privacy related questions. Annotation process consists of collecting information about various privacy sensitive regions, including primary and secondary privacy regions, and additional personal information for one chosen person and their recording in annotation file. More specifically, for each video sequence in our experimental dataset, one person was chosen and his/her body silhouette was annotated manually using ViPER-GT annotation tool. All annotations were recorded and stored in XML format for every frame where the region is visible.

---

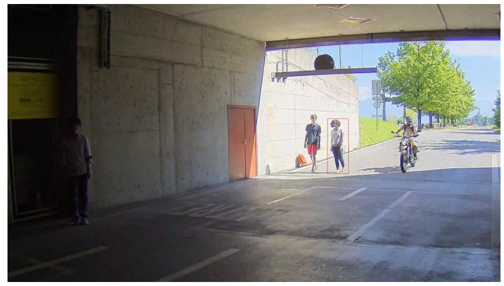
*`http://viper-toolkit.sourceforge.net`

††`http://www.adobe.com/devnet/cinemadng.html`

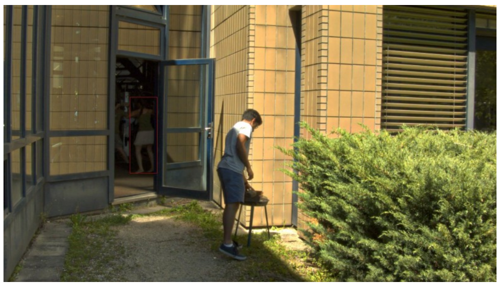
‡‡`http://www.cybercom.net/~dcoffin/dcraw/`



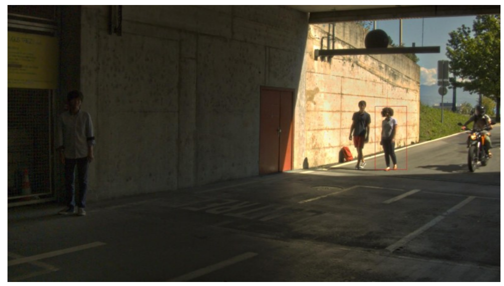
(a) Bosch camera - Entrance1



(b) Bosch camera - Parking2



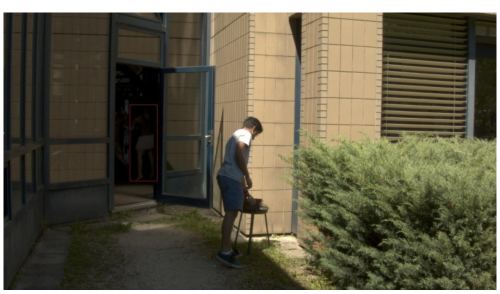
(c) BMPCC HDR camera - Entrance1



(d) BMPCC HDR camera - Parking2

(e) JAI camera - Entrance1

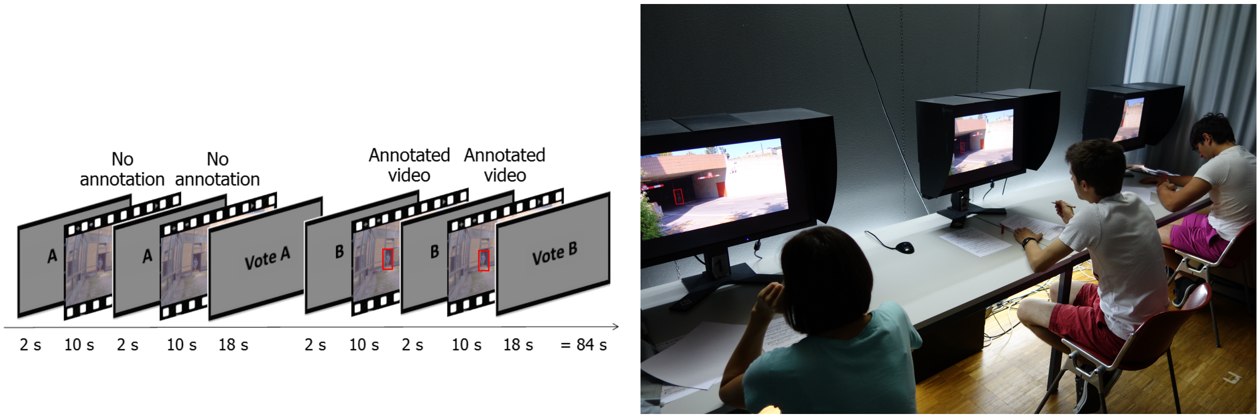
(f) JAI camera - Parking2



(g) BMPCC LDR camera - Entrance1

(h) BMPCC LDR camera - Parking2

Figure 2: Examples of two sequences used in the experiment representing the bright-to-dark (left column) and dark-to-bright (right column) scenario. (a) - (f) show HDR versions of the content recorded by all HDR acquisition systems and (g) and (h) LDR version.



(a) (b)

Figure 3: (a)Order of presented video test stimuli. (b)Testing room.

4.2 Subjective evaluation methodology

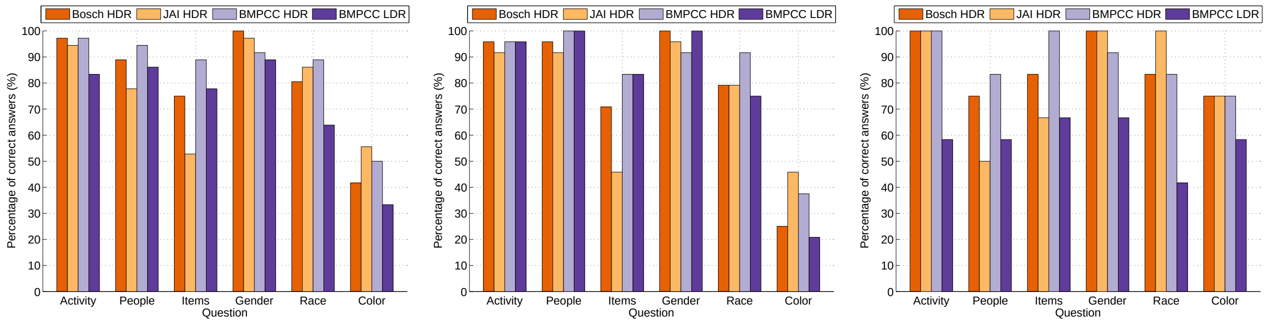
To evaluate privacy intrusiveness of HDR video data in our dataset, we asked subjects the following six questions:

1. Which of the following activities (actions) is happening in the video?
2. How many people do you see in the video?
3. Is there any of the following items in the video (you can select several)?
4. What is the GENDER of the person inside red box?
5. What is the RACE of the person inside red box?
6. What is the COLOR of upper body clothes (e.g. T-shirt) of the person inside red box?

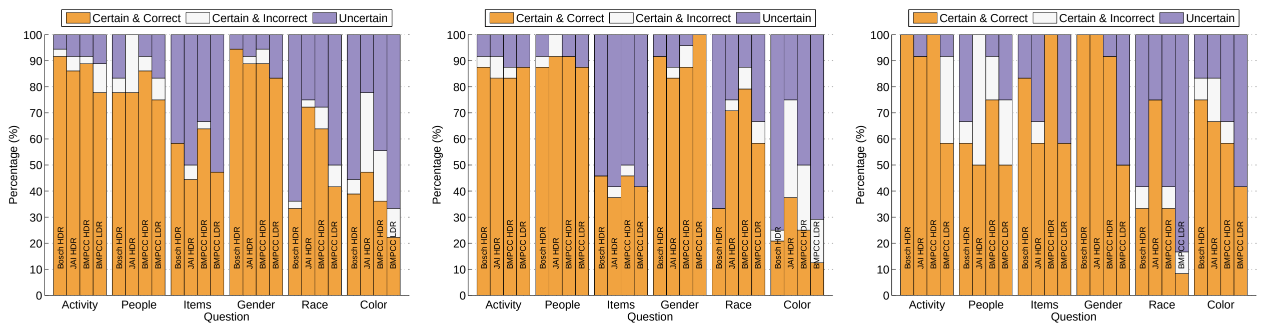
After each of these questions, subjects were also asked to evaluate how certain they are about their answer on a five point scale including labels *Very Sure*, *Sure*, *Neutral*, *Not So Sure*, and *Unsure*.

Four acquisition systems were compared in the subjective experiments, three HDR cameras and an LDR camera, video from which was simulated by using output of BMPCC camera (see Section 4.1 for details). It resulted in $4 \times 6 = 24$ video sequences, which were split into four individual subsets in such a way that each subset included only video sequences recorded by one acquisition system. Therefore, the test was divided into four individual sessions in which subjects evaluated only one acquisition system, i.e. six different video sequences. The order of the stimuli was played to subjects according to Figure 3(a). First, the video stimuli without annotation was shown twice, followed by answering first set of questions related to content intelligibility (questions 1-3). Second, the same video stimuli with an annotation was shown and questions related to privacy (question 4-6) were asked afterwards. Special care was taken to make sure that subjects are not asked for the same person in two video stimuli or that subjects see investigated person in previous stimuli.

The test was conducted in eight different time slots, each occupied by three subjects (see setup of the testing room in Figure 3(b)). The sequences were displayed on professional EIZO monitors CG301W in a testing environment complying with the ITU recommendation BT.500-13.[25] A total of 24 naïve people took part in the test campaign. Three (12.5%) of the observers were female and the age of the subjects ranged from 19 to 35 years old, with a median of 25.5 years old. All participants were screened for correct visual acuity and color vision using Snellen charts and Ishihara charts respectively. The training of the subjects of each group was conducted before the first test session, where oral instructions were provided to explain the task and a viewing session was performed to allow the subject to familiarize with the assessment procedure. To collect evaluation scores, subjects were provided with scoring sheets to enter their scores. The scores were then offline converted into electronic version. All the scores were converted by one operator and crosschecked by a second.



(a) All six sequences (b) Four bright-to-dark sequences (c) Two dark-to-bright sequences



(d) All six sequences (e) Four bright-to-dark sequences (f) Two dark-to-bright sequences

Figure 4: Evaluation results obtained by averaging subjective scores across all subjects. (a) Percentage of correct answers of each question for each acquisition system; (b) Percentage of correct answers of each question for each acquisition system (bright-to-dark); (c) Percentage of correct answers of each question for each acquisition system (dark-to-bright); (d) Percentage of three different types of answers of each question for each acquisition system; (e) Percentage of three different types of answers of each question for each acquisition system (bright-to-dark); (f) Percentage of three different types of answers of each question for each acquisition system (dark-to-bright).

4.3 Results

The results of subjective evaluation are shown in Figure 4. Figure 4(a) illustrates the percentage of correct answers of each question and each acquisition system (Bosch HDR, JAI HDR, BMPCC HDR and BMPCC LDR), for all six sequences. The figure compares three different HDR acquisition systems and one simulated LDR video. For each question, BMPCC HDR video always resulted in higher rate of correct answers than BMPCC LDR video, since HDR videos can provide more detail information for observers, e.g. something that hides in deep dark or over-saturated bright areas. Regarding the difference between BMPCC LDR and other two HDR acquisition systems, Bosch and JAI also provide more correct answers than BMPCC LDR in most questions but with some exception, for example, JAI HDR in questions 2 and 3 (*People* and *Items*), and Bosch HDR in question 3 (*Items*) have even less correct answers than BMPCC LDR. The reason is that both JAI and Bosch cameras have Field of View (FoV) larger than BMPCC camera so the people or items in the video shot by JAI or Bosch camera appear to be relatively smaller in comparison to video sequences acquired with BMPCC. In addition, video sequences shot by JAI camera have a lower resolution (1290×966 pixels) in comparison to other two acquisition systems Bosch and BMPCC (1920×1080 pixels). Above mentioned differences could negatively influence the judgment on how many people and what item(s) exist in the video recorded by JAI and Bosch camera. Mutual comparison of Bosch HDR and BMPCC HDR indicates a slightly better performance for BMPCC camera. Whereas for questions 2, 3, 5 and 6 (*People*, *Items*, *Race* and *Color*), BMPCC HDR gives more correct answers, for other two questions it has less or equal correct answers.

Figure 4(d) provides further analysis to compare HDR and LDR video as well as different HDR acquisition systems, which takes into account the certainty of answers. Since each question was followed by a sub-question of how sure the observer was about his/her answer, we can use the answer of this sub-question to measure the certainty of their judgment. In terms of correctness and certainty, we classified all answers into the following three categories:

Certain & Correct Correct answer with *Very sure* or *Sure* in certainty sub-question.
Certain & Incorrect Incorrect answer with *Very sure* or *Sure* in certainty sub-question.
Uncertain *I don't know* answer or any answer with *Not so sure*, *Unsure* or *Neutral* in certainty sub-question.

Figure 4(d) shows the distribution of the three types of answers to each question and each acquisition system for all six sequences. In general, all three versions of HDR video have higher rate of Certain & Correct answers and lower rate of Uncertain answers than BMPCC LDR video. An exception is that JAI HDR and Bosch HDR also have high uncertainty in questions 2 and 3 (*People* and *Items*), which is due to the same reason of larger FoV and lower resolution (JAI) as explained above. In terms of Certain & Incorrect answers, results may look confusing. For questions such as 1 and 5 (*Activity* and *Race*), BMPCC LDR generated the most Certain & Incorrect answers but in other questions such as 3 and 4 (*Items* and *Gender*) it had no Certain & Incorrect answer at all. Besides that, there is also no clear winner between the three different HDR acquisitions in terms of Certain & Incorrect answers. The main reason for this is that people can make wrong judgment when the video content itself is confusing and error-prone, however, higher video quality (HDR) made people even sure about their wrong answers.

Further analysis was made by focusing on different scenarios of the evaluated video. Since the answers also depend on the scenario and content of the video sequences, it is interesting to understand how different scenarios affect subjective judgment. As described in Section 4 and illustrated by Figure 2, for each acquisition system, four of the six video sequences are bright-to-dark scenarios (monitoring bright areas) and two are dark-to-bright scenarios (monitoring dark areas). Results for the two scenarios are shown in the second and third column of Figure 4 respectively. From Figure 4(b), percentage of correct answers for bright-to-dark scenario's LDR video is not as low as the case in the average situation, which Figure 4(a) shows. From certainty distribution of bright-to-dark scenario (Figure 4(e)), there is no significant difference between LDR and three HDR versions either. However, if we check the results of dark-to-bright scenario's LDR video in Figure 4(c) and Figure 4(f), LDR is clearly worse than other HDR video sequences for almost every question in terms of both correctness and uncertainty. This is because in the LDR video of bright-to-dark scenarios, when we monitor activities in bright area, only highlights containing the main information of HDR scene are preserved. Due to the capabilities of used acquisition systems and their limitation in terms of noise and shutter speed, the dark areas (shadows) don't contain useful information and thus the difference between HDR and LDR is negligible. On the other hand, LDR video of dark-to-bright scenarios preserves only the shadows of HDR scene and thus bright areas are over-saturated. Then it made it very hard for people to see the activity happening in bright areas. While, HDR video of dark-to-bright scenario kept highlight part so that it made big difference with its LDR version.

Another thing we should notice from the results is that some questions generated more incorrect or uncertain answers than the others, e.g. question 2 and 6 (*Items* and *Color*). Question 2 (*Items*) highly depends on where people focus their attention. People might not notice some small items lying on a corner where there was nothing happening. Besides, the video used for test was not calibrated with a unique white balance so some colors might be very ambiguous like dark green with brown or dark blue with gray.

To summarize, the tone-mapped HDR video shows more information including privacy details than LDR video, as the number of correct answers is higher and uncertainty is lower. However, the superiority of HDR video is scenario dependent, which relied on the target and content of video surveillance, e.g. what and where we want to monitor. In addition, it is hard to tell which of the three different HDR acquisition systems is the best because they have different angles of view and resolution. While in general, BMPCC HDR is slightly better than the other two because it has the smallest angle of view and HD resolution.

5. CONCLUSION

This paper describes the creation and generation of an HDR dataset of moving pictures suitable for evaluation of visual privacy protection tools. The HDR video dataset consists of 18 indoor, 3 night and 47 outdoor video sequences recorded by three different HDR cameras. Created HDR dataset is available as an extension part of PEViD public database.

To demonstrate the usability of the created PEViD-HDR video dataset, a preliminary subjective experiment was conducted to compare video recorded by three HDR acquisition systems and one simulated LDR system. The evaluation results show that HDR sequences are more privacy intrusive in comparison to LDR content, since HDR video contains more visible detailed information. These results indicate that HDR technology and HDR video data is more challenging in terms of privacy intrusiveness and the development of privacy protection tools.

ACKNOWLEDGMENTS

This work has been conducted in the framework of the EC funded Network of Excellence VideoSense, Swiss SERI project Compression and Evaluation of High Dynamic Range Image and Video, COST IC1005 The digital capture, storage, transmission and display of real-world lighting HDRi, and COST IC1003 European Network on Quality of Experience in Multimedia Systems and Services QUALINET. This work was also partially supported by the project COST CZ LD12018 Modeling and verification of methods for Quality of Experience (QoE) assessment in multimedia systems - MOVERIQ of the Ministry of Education, Youth and Sports of the Czech Republic, and by the grant No. P102/10/1320 Research and modeling of advanced methods of image quality evaluation of the Czech Science Foundation. The authors would like to thank the actors and the participants in subjective experiment for their assistance. We also thank PCS Ltd., the official Bosch Security Systems distributor in the Czech Republic, for their cooperation.

REFERENCES

[1] Mantiuk, R., Cichowicz, M., and Smyk, M., "Implementation of HDR photographic pipeline in mobile devices," in [*Proceedings of the 9th International Conference on Image Analysis and Recognition - Volume Part I*], *ICIAR'12*, 367–374, Springer-Verlag, Berlin, Heidelberg (2012).

[2] Korshunov, P. and Ebrahimi, T., "PEViD: privacy evaluation video dataset," in [*SPIE Applications of Digital Image Processing XXXVI*], **8856** (Aug. 2013).

[3] Korshunov, P., Nemoto, H., Skodras, A., and Ebrahimi, T., "Crowdsourcing-based evaluation of privacy in HDR images," in [*SPIE Photonics Europe 2014, Optics, Photonics and Digital Technologies for Multimedia Applications*], (Apr. 2014).

[4] Ebrahimi, T., Abdeljaoued, Y., Ventura, R. F. I., and Escoda, O. D., "MPEG-7 camera," in [*Proceedings of International Conference on Image Processing (ICIP 2001)*], **3**, 600 –603 (2001).

[5] Senior, A., Pankanti, S., Hampapur, A., Brown, L., Tian, Y.-L., Ekin, A., Connell, J., Shu, C., and Lu, M., "Enabling video privacy through computer vision," *IEEE Security and Privacy* **3**, 50–57 (May 2005).

[6] Park, S.-W., Han, J. W., and Shin, S.-U., "Secure service mechanism of video surveillance system based on h.264/svc," in [*International Conference on Information Technology and Multimedia (ICIM 2011)*], 1 –4 (Nov 2011).

[7] Aved, A. J. and Hua, K. A., "A general framework for managing and processing live video data with privacy protection," *Multimedia Syst.* **18**(2), 123–143 (2012).

[8] Zhang, W., Cheung, S., and Chen, M., "Hiding privacy information in video surveillance system," in [*Proc. IEEE International Conference on Image Processing*], (Sep 2005).

[9] Dufaux, F. and Ebrahimi, T., "Scrambling for privacy protection in video surveillance systems," *IEEE Trans. on Circuits and Systems for Video Technology* **18**, 1168–1174 (Aug 2008).

[10] Wickramasuriya, J., Datt, M., Mehrotra, S., and Venkatasubramanian, N., "Privacy protecting data collection in media spaces," in [*Proceedings of the 12th annual ACM international conference on Multimedia*], 48–55 (Oct. 2004).

[11] Korshunov, P. and Ebrahimi, T., "Using warping for privacy protection in video surveillance," in [*18th International Conference on Digital Signal Processing (DSP)*], 1–6 (2013).

[12] Korshunov, P. and Ebrahimi, T., "Using face morphing to protect privacy," in [*IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS)*], 208 – 213 (Aug. 2013).

[13] Newton, E., Sweeney, L., and Malin, B., "Preserving privacy by de-identifying face images," *IEEE Trans. on Knowledge and Data Engineering* **17**, 232–243 (Feb 2005).

[14] Korshunov, P. and Ooi, W. T., "Video quality for face detection, recognition, and tracking," *ACM Trans. Multimedia Comput. Commun. Appl.* **7**, 14:1–14:21 (Sept. 2011).

[15] Dufaux, F. and Ebrahimi, T., "A framework for the validation of privacy protection solutions in video surveillance," in [*Proceedings of IEEE International Conference on Multimedia & Expo (ICME 2010)*], (July 2010).

[16] Korshunov, P., Melle, A., Dugelay, J.-L., and Ebrahimi, T., "A framework for objective evaluation of privacy filters in video surveillance," in [*SPIE Applications of Digital Image Processing XXXVI*], **8856** (Aug. 2013).

[17] Korshunov, P., Araimo, C., De Simone, F., Velardo, C., Dugelay, J., and Ebrahimi, T., "Subjective study of privacy filters in video surveillance," in [*2012 IEEE 14th International Workshop on Multimedia Signal Processing (MMSP)*], 378–382 (Sept. 2012).

[18] Ledda, P., Chalmers, A., Troscianko, T., and Seetzen, H., "Evaluation of tone mapping operators using a high dynamic range display," in [*ACM SIGGRAPH 2005 Papers*], *SIGGRAPH'05*, 640–648, ACM, New York, NY, USA (Aug. 2005).

[19] Yoshida, A., Blanz, V., Myszkowski, K., and Seidel, H.-P., "Perceptual evaluation of tone mapping operators with real-world scenes," **5666**, 192–203, Human Vision and Electronic Imaging X, IS&T/SPIE's 17th Annual Symposium on Electronic Imaging (2005).

[20] Park, S. H. and Montag, E. D., "Evaluating tone mapping algorithms for rendering non-pictorial (scientific) high-dynamic-range images," *J. Vis. Comun. Image Represent.* **18**(5), 415–428 (2007).

[21] Kuang, J., Yamaguchi, H., Liu, C., Johnson, G. M., and Fairchild, M. D., "Evaluating HDR rendering algorithms," *ACM Trans. Appl. Percept.* **4**, 9:1–9:27 (July 2007).

[22] Annighöfer, B., Tajbakhsh, T., and Grigat, R.-R., "Prediction of results from subjective evaluation of real-time-capable tone-mapping operators applied to limited high-dynamic-range images," *Journal of Electronic Imaging* **19**, 011015–011015–12 (Jan. 2010).

[23] Mai, Z., Doutre, C., Nasiopoulos, P., and Ward, R., "Subjective evaluation of tone-mapping methods on 3D images," in [*17th International Conference on Digital Signal Processing (DSP)*], 1–6 (July 2011).

[24] Kuang, J., Johnson, G., and Fairchild, M., "iCAM06: A refined image appearance model for hdr image rendering," *Journal of Visual Communication and Image Representation* **18**(5), 406–414 (2007).

[25] ITU-R BT.500-13, "Methodology for the subjective assessment of the quality of television pictures." International Telecommunication Union (January 2012).